Postal Regulatory Commission
Submitted 5/25/2012 11:12:55 AM
Filing ID: 82755
Accepted 5/25/2012

BEFORE THE
POSTAL REGULATORY COMMISSION
WASHINGTON DC 20268-0001

| | |
|---|---|
| Priorities for Future Data Collection and Analytical Work Relating to Periodic Reporting | Docket No. RM2011-3 |

Scoping Study Report of the United States Postal Service

## TABLE OF CONTENTS

## I. INTRODUCTION

Volume variable city carrier street time product costs are calculated in three steps. First, total accrued costs are divided into activity cost pools. Second, variabilities are applied to each of the cost pools to find their volume variable costs. Third, the volume variable costs are distributed to the products handled in the cost pool activity. In the past, both the cost pools and the variabilities were estimated through special studies, and the variable costs were distributed on the basis of City Carrier Cost System (CCCS) data.

This scoping study is the first step in investigating an efficient and effective way of updating the first two steps: forming the cost pools and estimating the variabilities. We reviewed important changes in the delivery environment that have occurred since the 2002 City Carrier Street Time Study (CCSTS), investigated ongoing databases to discover their potential usefulness in estimating cost pools and variabilities, and explored model specification choices including options for both forming cost pools and estimating variabilities. We report on our findings in this report.

The next section describes our findings on important operational changes in the street time environment that could affect subsequent modeling efforts. This is followed by a section describing, in detail, our findings from reviewing the various ongoing datasets that are currently available and cover city carrier street activities. The next two sections describe our efforts in model selection and estimation, first in the area of construction of the cost pools and then in the area of variability estimation. We conclude with a brief enumeration of our main findings.

## II. CHANGES IN THE DELIVERY ENVIRONMENT SINCE THE LAST STUDY

There have been a number of important operational changes since the last study which could affect the way street time costs are incurred. The operational changes are listed below. In each case, we describe the operational change and then discuss the implications for city carrier street time costing.

- Use of Delivery Point Sequencing (DPS) is widespread. Nearly all routes have a DPS tray or bundle to take to the street and DPS volumes are the largest volume grouping that is taken to the street. This suggests that future disaggregated delivery time models should include a separate variable for DPS.

- Use of the Flats Sequencing System (FSS) is expanding but flats sequencing is performed for only a minority of ZIP Codes. When the FSS is in use, it creates an additional container or bundle of mail that must be handled by the carrier. This is not a problem for motorized routes, but can cause additional work for foot and park and loop routes. This suggests that future delivery time models should explicitly account for the street time caused by FSS when it is present, perhaps as including FSS mail as a separate variable.

- Carriers are spending a higher proportion of their day on the street. This is due to the success of DPS in reducing casing time and the decline in volume. Carriers routinely spend six hours on the street and are moving toward seven or more with FSS implementation. This fact should be used in interpreting any changes in cost pools, productivities or variabilities.

- The Postal Service has successfully used its Carrier Optimal Routing (COR) system to reduce vehicle travel time both in terms of driving to and from routes and driving within routes. This fact should be used in interpreting any observed changes in cost pools.

- Every Door Direct Mail (EDDM) unaddressed mail pieces are handled on the street as sequenced mailings. That suggests they could be included in delivery time models in the same way as sequenced mailings.

- Management of carrier delivery occurs at the ZIP Code level, not the individual route level. While routes are not exactly eight hours, variation in daily route time is limited as routes are reconfigured to match ZIP Code volume and delivery points. This suggests that estimating of delivery time variabilities should be done at the ZIP Code level.

- Pivoting is used much more extensively to cover routes. This is in part because of the reduction in volume and in part because of reduced office time. Increased pivoting suggest that the route is a less useful unit of observation for measuring carrier costs and reflects the fact that carrier street time costs are managed at the ZIP Code level.

- Cased letters and flats are considered “residual mail” and are combined into one bundle for street delivery. This suggests that future disaggregated delivery time models should include just one variable for all cased mail.

- Carrier pickups are still very small, but when they occur they could cause a deviation by the carrier. They are sufficiently infrequent, however, so as not to add a material amount of street time.

- Since FY2002, there has been a 14 percent reduction in the number of city routes. This reduction has not been balanced by route type, as the Postal Service has cut walking routes (foot and park and loop) by higher percentages than the overall average. This suggests it may be important to take into account the types of routes being used when estimating street time costs.

Route Type Distribution

| | FY2002 | FY2006 | FY2010 | FY2012 | % Change from FY 2002 |
|---|---|---|---|---|---|
| Curb | 39,231 | 38,884 | 37,276 | 36,510 | -6.9% |
| Dismount | 25,210 | 25,752 | 23,930 | 23,607 | -6.4% |
| Foot | 12,658 | 10,929 | 9,134 | 8,226 | -35.0% |
| Park & Loop | 89,263 | 87,157 | 80,795 | 75,240 | -15.7% |
| Other | 698 | 735 | 727 | 330 | -52.7% |
| Total | 167,060 | 163,457 | 151,862 | 143,913 | -13.9% |

*Source: City Carrier Cost System Frame*

## III. REVIEW OF ONGOING DATASETS AND THEIR USEFULNESS IN COST ESTIMATION.

One of the primary tasks of the scoping study was to investigate existing datasets that might hold potential for updating city carrier street time costs without the need for an extensive field study. To that end, we surveyed operational and data experts within the Postal Service to identify possible sources of city carrier data. We identified five different data sets that relate to city carrier street time activities:

- The Form 3999 Dataset
- The City Carrier Cost System (CCCS) Dataset
- Carrier Optimal Routing (COR) System Dataset
- The Delivery Operations Information System (DOIS) Daily Dataset
- The Managed Service Point (MSP) Scan Dataset

We then investigated each of these datasets to evaluate its potential for either updating the city carrier street time cost pools or for estimating city carrier street time variabilities. Our evaluation of each dataset was based upon the following criteria:

| | |
|---|---|
| Amount: | How much data are available from the operational database? Are there sufficient data available for the required analysis? |
| Completeness: | Is the dataset relatively complete? Are there many missing observations? |
| Consistency: | Are the data relatively consistent throughout the dataset? Are descriptive statistics reasonably similar across subsets? |
| Correctness: | Do the data appear to be correctly recorded? To what degree do the data contain clear errors? |

| | |
|---|---|
| Timeliness: | When were the data collected? Are the data of acceptable vintage? |
| Integrity: | What processes were used to collect the data? Are there any protocols to ensure data quality? How are the data used in their primary process? |
| Usefulness: | How applicable are the data for the task at hand? Can they be used in an effective way? |

A detailed description of the results of our analysis for each of the five datasets we evaluated is presented in this section. Before the detailed results are presented, we summarize our findings about each dataset.

- The Form 3999 data appears to be potentially useful for cost estimation. It provides detailed information regarding carrier activities for about 140,000 routes. This would seem to be more than sufficient for forming cost pools. In addition it provides activity times and delivered volumes for over 10,000 ZIP Codes which would seem to be enough data for possibly estimating street time variabilities.

- In addition to its use in forming the distribution key for volume variable street time costs, the City Carrier Cost System (CCCS) database is useful for forming a benchmark for evaluating DOIS data. However, because it is not collected at the ZIP Code level, the CCCS dataset is not useful in estimating street time variabilities.

- The Carrier Optimal Routing (COR) system is a management tool that creates efficient routes but does not provide any additional information that is not already in the route inspection (Form 3999) data. In addition, COR data is not useful for estimating variabilities, because city carrier street costs are managed by ZIP Code not block-face, and it does not provide any additional volume data at the ZIP Code level.

- The Delivery Operations Information System (DOIS) daily data cover nearly all of the routes and ZIP Codes in the country and provide volume and time data for every delivery day in the year. The data set is extensive, timely, and acceptably complete. A comparison of DOIS daily data with CCCS data shows that the DOIS daily data are reasonably accurate and consistent. The DOIS daily data would appear to hold the potential for estimating street time variabilities.

- The Managed Service Point (MSP) program uses mobile data collection devices to scan barcodes placed strategically along the city carrier's line of travel. A series of barcodes are located in the office and on the route so that the supervisor can follow a carrier’s progress during the day. It is a management tool that gives supervisors knowledge of a carrier’s general location on the route and is useful to the Postal Service in managing street time. For example, it allows the supervisor to monitor the consistency of delivery time. However, MSP data does not provide the necessary information about specific carrier activities to aid in calculating attributable costs by product.

## A. The Form 3999 Dataset

### 1. Data Set Description

The Form 3999 data set consists of one observation for each city carrier route in the country. The data come from the day on which the route is evaluated. A route evaluation is the process through which the Postal Service collects data on the times the carrier spends in the various office (Form 1838) and street (Form 3999) activities on a route. The Postal Service also collects data on the volumes delivered by that carrier.

The Form 3999 database includes the most recent route evaluations for 140,794 city carrier routes. These evaluations occurred over the period from 2006 through 2012 but, as the following table shows, 99 percent of the evaluations occured in the most recent four years. The evaluations therefore took place as the Postal Service downsized its carrier network in response to declines in volume and widespread adoption of DPS.

| | Route Evaluations | Percentage | Route Evaluations with Positive Delivery Time | Percentage |
|---|---|---|---|---|
| 2006 | 36 | 0.03% | 36 | 0.03% |
| 2007 | 17 | 0.01% | 17 | 0.01% |
| 2008 | 69 | 0.05% | 68 | 0.05% |
| 2009 | 3,377 | 2.40% | 3,371 | 2.40% |
| 2010 | 24,953 | 17.72% | 24,915 | 17.73% |
| 2011 | 96,293 | 68.39% | 96,087 | 68.39% |
| 2012 | 16,049 | 11.40% | 16,007 | 11.39% |

The route evaluation process includes recording the times that the carrier is engaged in the various office and street activities and a mail count conducted by the delivery unit manager or designee. This process includes unannounced selective checks on all of the routes being inspected to verify the accuracy of the mail count. In addition a route examiner makes a physical inspection of the route and then accompanies the carrier for the full tour on the day of the inspection.

2. Key Variables

The Form 3999 dataset is a cross-sectional database in which the unit of observation is the route. Each observation contains the activity times and volumes delivered on that route during its route evaluation. The Form 3999 thus provides a profile of the daily activity for each city carrier route in the country. Also, because the data set covers all routes, the data can be aggregated by ZIP code. Key variables include:

| | |
|---|---|
| Travel Within Time | Delivered FSS Pieces |
| Sector Segment Time | Delivered Deviation Parcel Pieces |
| Collection from SLB Time | Delivered Sequenced Pieces |
| Relay Time | Parcel Delivery Time |
| Delivered DPS pieces | Accountable Delivery Time |
| Delivered Cased Flat pieces | Possible Deliveries |
| Delivered Cased Letter Pieces | Route Delivery Type |

Average values for various Form 3999 variables can provide a daily profile for city carrier routes. As the nature of city carrier activity has changed, the amount of time spent by carriers on the street has increased. That change is illustrated in the following table which shows, on average, city carriers spend nearly five hours in "sector segment" or delivery time. This is time spent within the delivery sections on the route. In addition, the carrier has to drive to and from the route, drive within the route, deliver parcels and accountables and undertake relays.

Form 3999 Average Values for Delivery Variables*

| **Variable** | **Mean** |
|---|---|
| Sector Segment Hours | 4.49 |
| Travel To and From Hours | 0.27 |
| Travel Within Hours | 0.17 |
| Parcel/Accountable Hours | 0.23 |
| Collection from SLB Hours | 0.01 |
| Relay Hours | 0.22 |
| Possible Deliveries | 610.3 |
| Volume | 2,514.6 |
| Number of Observations | 140,501 |

* *Means are calculated on the routes that reported positive sector segment hours*

3. <u>Potential Uses</u>

The Form 3999 data set contains the times required to accomplish a set of street activities for virtually all city carrier routes. These times can be cumulated and used to find the proportions of time, across the country, for each of the various activities performed on the street. That makes the Form 3999 data a potentially important source for cost pool formation. Cost pools are constructed by multiplying current total accrued street time by the time proportions for the various street activities. The Form 3999 data could be used to estimate those time proportions.

In addition, because it can be aggregated to the ZIP Code and because it includes delivered volumes as well as recorded street time for delivery activities, the Form 3999 data could possibly be used for estimating street time variabilities.

4. Evaluation of the Usefulness of the Dataset for Cost Estimation

The Form 3999 data appears to be potentially useful for cost estimation. It provides about 140,000 observations on route activities. This would seem to be more than sufficient for forming cost pools. In addition, it provides activity times and delivered volumes for over 10,000 ZIP Codes which would appear to be enough data for estimating street time variabilities. The Form 3999 data are complete in terms of covering the national network of city carriers. Finally, the recorded data cover the entire time spent on the street. However, the data set does not include accountables or collection volume.

Examination of the data by route type provides evidence of consistency and reasonableness. As the next table shows, motorized routes spend more time in delivery activities and less time in street support activities such as travelling to and from the route and relay. In addition, because they are motorized, both curbline and dismount routes handle a larger amount of volume and serve more delivery points on a daily basis than do foot and park and loop routes.

Daily Average Values by Route Type*

| Type of Route | Sector Segment Time (hours) | Travel To/From Time (hours) | Relay Time (hours) | Volume (pieces) | Possible Deliveries | Number of Routes |
|---|---|---|---|---|---|---|
| Curbline | 4.63 | 0.28 | 0.06 | 2,945.8 | 653.7 | 36,007 |
| Dismount | 4.56 | 0.28 | 0.09 | 2,773.6 | 657.9 | 23,295 |
| Foot | 3.76 | 0.37 | 0.23 | 2,420.4 | 627.2 | 7,824 |
| Park and Loop | 4.48 | 0.25 | 0.33 | 2,232.7 | 563.3 | 73,016 |
| Other | 2.77 | 0.20 | 0.12 | 1,879.2 | 387.4 | 339 |

** Note that 20 routes did not have a recorded route type. The first two route types, curbline and dismount are considered to be "motorized" routes.*

The Form 3999 data set contains the expected number of "unusual observations" or "outliers," but there appear to very few clearly erroneous observations. Moreover because the data are used to construct routes, care is taken avoid errors.

Finally, while it would be ideal to have data all from the current year, over 99 percent of the evaluations were performed in the most recent four years. An examination of the data shows that they are consistent through time. As the following table shows, the street time proportions have been stable over the period during which the route evaluations have been performed. Moreover, consistent with Postal Service procedures, including the use of COR, there has been a modest decline in travel time and a modest increase in delivery time.

Street Time Proportions Based on Form 3999 Data

| | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|
| Sector Segment Delivery | 83.0% | 83.3% | 83.5% | 83.5% |
| Parcel/Accountable Delivery | 3.7% | 4.0% | 4.3% | 4.8% |
| Relay | 4.2% | 4.0% | 4.1% | 3.7% |
| Travel To/From Route | 5.6% | 5.2% | 4.9% | 5.1% |
| Travel Within | 3.4% | 3.3% | 3.1% | 2.6% |
| Collections From SLB | 0.1% | 0.1% | 0.2% | 0.2% |
| Number of Observations | 3,371 | 24,915 | 96,087 | 16,007 |

## B. The City Carrier Cost System (CCCS) Dataset

### 1. Data Set Description

The City Carrier Cost System (CCCS) is an ongoing statistical system that estimates annual delivered and collected volume by product on city letter routes. CCCS utilizes a stratified design and samples more than 8,000 route days annually.

The CCCS is a continuous, ongoing cross-sectional statistical study, or probability sample of city carrier route-days. As mentioned above, more than 8,000 CCCS samples are scheduled each Fiscal Year. For each selected route-day, a sample of mail is selected, and for each selected mail piece, the class, product, and other characteristics are recorded directly into a portable microcomputer using the Computerized On-Site Data Entry Systems (CODES) software. CCCS estimates are used to distribute attributable city letter route street time costs to products within each cost pool.

2. Key Variables

CCCS provides detailed estimates, by product and shape, of the volume of mail delivered by city carriers. Examples of key variables, by shape, include DPS letters, cased letters, cased flats, deviation parcels, FSS flats, sequenced volume, and collected volume.

3. Potential Uses

The CCCS is already very useful in providing the distribution key, by product, for city carrier street time costs. However, given that CCCS provides detailed information by shape and that the data is carefully collected and evaluated, it holds the potential for other uses. First, it can provide a benchmark for evaluating DOIS volume data. Second, because it provides volumes by product, it could possibly be matched with DOIS hours to estimate a variability equation.

4. Evaluation of the Usefulness of Data Set for Cost Estimation

In general, each route sampled by CCCS is only chosen once per year. In addition CCCS does not test all routes in a year. This means that it cannot be aggregated to the ZIP Code level in a way which provides all or nearly all routes for each ZIP Code. As a result, its data are not useful for variability analysis. However, CCCS data are required to accurately assign attributable costs to products.

The CCCS volume data, which are carefully collected by trained personnel, can be compared with corresponding DOIS data. DOIS data relies on machine counts (DPS and FSS), linear measurements (cased letters and flats), residential deliveries (sequenced mail), and piece counts (parcels). The results from the comparison of CCCS and DOIS data suggest that DOIS volume data are appear acceptable for possibly updating the regular delivery variability estimated from 2002 City Carrier Street Time Study (CCSTS). The discrepancies between CCCS and DOIS volume were generally small and application of both data sets to regression analysis produced similar results.[1]

In addition to its use in forming the distribution key for volume variable street time costs, the CCCS database is useful for forming a benchmark for evaluating DOIS data. However, because it is not collected at the ZIP Code level the CCCS dataset is not useful in estimating street time variabilities.

## C. The Carrier Optimal Routing (COR) System Dataset

### 1. Data Set Description

The Carrier Optimal Routing (COR) system uses geographic information from mapping software and route information supplied through DOIS to realign territory within a delivery zone and adjust lines of travel on routes. Using COR, the Postal Service attempts to optimize a zone by eliminating unnecessary travel time (travel to, travel from, travel within) and relay time. The data used in COR includes times recorded in the route evaluation process (Form 3999 data) and volume data obtained through DOIS.

[1] The comparison is discussed in detail in the subsection describing the DOIS Daily dataset, below.

2. Key Variables

On a route-by-route basis, COR includes the times for street activities, like travel to and from a route and sector segment times. Where appropriate, these times are broken down by ZIP+4 sections of the route. In addition, DPS and FSS volumes are assigned to blockface (9 digit ZIP Codes) through the use of machine counts, and other route volumes are assigned to blockface in proportion to the DPS and FSS volume. Parcel volumes are not included.

3. Potential Uses

Because the COR data set includes measures of both activity times and volumes for all of the routes in a ZIP Code it holds the potential to serve as a basis for estimating variabilities for street activities. In addition, the activity times in COR could potentially be used to estimate cost pools.

4. Evaluation of the Usefulness of Data Set for Cost Estimation

Investigation of COR data indicated that it is not a useful source to use to update the 2002 CCSTS. The CCSTS serves two main functions, 1) divide the city letter carrier's street time into relevant cost pools and 2) calculate the cost impacts due to a change in volume. COR is not useful for either purpose.

COR is a management tool that creates efficient routes but does not provide any additional information that is not already in the route inspection (Form 3999) data. In addition, COR data is not useful for estimating variabilities, because city carrier street costs are managed by ZIP Code, not block-face, and it does not provide any additional volume data at the ZIP Code level.

## D. The Delivery Operations Information System (DOIS) Daily Dataset

### 1. Data Set Description

Part of the Delivery Operations Information System (DOIS) is the daily recording of street times, office times, and delivered volumes for virtually all city carrier letter routes in the country. Office and street hours for each route are recorded daily, and are taken from the Postal Service's Time and Attendance Control System (TACS). Automated mail volume piece counts are taken from End of Run (EOR) reports, and cased letter and flat volumes are recorded linearly and then converted to pieces. Finally, sequenced mailings, including both letters and flats, that do not require casing, are recorded as sets (either full or partial) and then converted to pieces based upon the number of residential deliveries on the route.

### 2. Key Variables

DOIS records machine counts of DPS and FSS volumes, linear measurements of cased letters and cased flats, a delivery point count of sequenced volume and a count of parcel volumes. It does not include measurement of collection volumes or accountables.

### 3. Potential Uses

Because it records delivered volumes for all routes in each ZIP Code, DOIS daily data holds the potential for providing the volume variables required for estimating variabilities for street time activities. In addition, because it records total street time for each route on a daily basis, DOIS daily data holds the potential for estimating overall street time variability.

### 4. Evaluation of the Usefulness of Data Set for Cost Estimation

There are number of characteristics of DOIS daily data that recommend it for use in variability estimation. First, it is plentiful and ongoing. It provides measures for street time and volumes delivered for virtually all ZIP Codes on a daily basis. Moreover, the data set is generally complete across ZIP Codes and covers most of the variables of interest. Finally, because the data are collected on an ongoing basis, there are no issues with timeliness. By its nature, a DOIS daily dataset is up to date.

One issue that has arisen in past examinations of DOIS data is the correctness of the data. While it is generally believed that machine counts for DPS and FSS are acceptably accurate, the use of linear measures (and piece conversions) to estimate cased mail raises a concern about the accuracy of this volume measure. One way of investigating this issue is to compare DOIS daily data with a data set of known accuracy. We did such a comparison between DOIS Daily volumes and CCCS volumes.

In FY2011, there were 8,288 route/days on which CCCS tests were taken. We matched 7,845 of those CCCS route days with their corresponding DOIS daily volume.

This means that the comparison data set has 7,845 observations, each of which includes CCCS measures of volume, DOIS measures of volume and DOIS measures of hours.

An initial comparison of the data sets can come from examining sample statistics and correlations across key variables in the data sets. In the first comparison, we examine total volumes and times per route.

| Variable | N | Mean | Std Dev | Minimum | Maximum |
|---|---|---|---|---|---|
| CCS Total Volume | 7,845 | 2,208 | 888.4 | 33 | 8,552 |
| DOIS Total Volume | 7,845 | 2,274 | 906.6 | 0 | 8,994 |
| Street Hours | 7,845 | 5.95 | 1.4 | 0 | 14.21 |
| Office Hours | 7,845 | 1.96 | 0.9 | 0 | 11.36 |

CCCS volumes are highly correlated with DOIS volumes across routes, but, as expected, both have a relatively low correlation with street hours. This is for two reasons. First, many street activities are not volume related, so street time has a relatively low volume variability. Second, a low correlation occurs at the route level because the Postal Service adjusts routes so that they are each approximately eight hours a day. Thus, one can observe both high and low volume routes with approximately the same amount of street hours.

Correlations between CCCS Volumes and DOIS Volumes and Times

| | CCCS Volume | DOIS Volume | Street Hours | Office Hours |
|---|---|---|---|---|
| CCCS Volume | 1 | 89.4% | 18.2% | 35.5% |
| DOIS Volume | | 1 | 16.0% | 40.8% |
| Street Hours | | | 1 | -13.8% |
| Office Hours | | | | 1 |

*All correlations are statistically significant.*

A more detailed level of comparison can be made by comparing the volumes by shape. In the current environment, carriers take the following containers or bundles to the street: DPS, Cased Mail, FSS, Sequenced, and Deviation Parcels.[2] The following table compares the average (daily) volumes per route both for CCCS and DOIS. The table shows that the volumes both in total and by "shape" are quite similar between DOIS and CCCS.

[2] The treatment of small parcels (SPRs) depends upon their size and shape. To the extent they can be, they are cased with flats. DOIS daily volume counts SPRs in with cased flats while CCCS counts them separately. For purposes of comparison, CCCS small parcels are combined with cased flats.

Comparison of Mean Volumes

| | **CCCS** | **DOIS** | **Difference (DOIS-CCCS)** |
|---|---|---|---|
| DPS | 1,447.9 | 1,489.7 | 41.8 |
| Cased Letters | 185.7 | 118.5 | -67.3 |
| Cased Flats &Parcels | 400.0 | 454.9 | 54.8 |
| Cased Mail | 585.8 | 573.3 | -12.4 |
| FSS | 34.4 | 36.9 | 2.5 |
| Sequenced | 128.8 | 161.3 | 32.5 |
| Dev. Parcels | 11.0 | 12.9 | 2.0 |
| Total | 2,208.2 | 2,274.2 | 65.9 |

Moreover, an argument can be made that the difference in daily volumes per route is overstated by using the difference in the means. To see this, one can compute the mean and median of the differences as opposed to just the difference of the means. In this approach the difference between the DOIS value for a particular volume category and the CCCS value for the same category is computed for each route. Then the distribution of those differences is examined. For example, the following table provides the distribution of differences in DPS volumes. The difference is calculated by subtracting the CCCS volume for a route from the DOIS volume for that same route.

Differences Between DOIS and CCCS Measured Volumes

| **DPS** | | | |
|---|---|---|---|
| Moments | | | |
| Mean | 41.8 | Std Dev | 205.7 |
| Median | 12.0 | Range | 7,052.0 |
| | | | |
| Quantile | | Estimate | |
| 100% | Max | 4332 | |
| 99% | | 731 | |
| 95% | | 205 | |
| 90% | | 111 | |
| 75% | Q3 | 40 | |
| 50% | Median | 12 | |
| 25% | Q1 | -1 | |
| 10% | | -15 | |
| 5% | | -35 | |
| 1% | | -200 | |
| 0% | Min | -2720 | |

This shows that the median difference is only 12 pieces, much smaller than the mean difference. The mean difference is inflated because of a small number of comparisons in which the DOIS volume is much larger than the CCCS volume. The table also presents the distribution of differences which shows that there is very little difference in measured volumes for the overwhelming majority of routes

The next table shows that the same pattern holds for cased mail. The median difference in volume is just four pieces.

Differences Between DOIS and CCCS Measured Volumes

| | | | |
|---|---|---|---|
| **Cased Mail** | | | |
| Moments | | | |
| Mean | -12.4 | Std Dev | 302.7 |
| Median | 4.0 | Range | 5,682.0 |
| | | | |
| Quantile | | Estimate | |
| 100% | Max | 2078 | |
| 99% | | 799 | |
| 95% | | 399 | |
| 90% | | 274 | |
| 75% | Q3 | 113 | |
| 50% | Median | 4 | |
| 25% | Q1 | -102 | |
| 10% | | -332 | |
| 5% | | -541 | |
| 1% | | -1003 | |
| 0% | Min | -3604 | |

The next two tables show that this pattern also holds for FSS volume and sequenced volume. Because neither of these volume types occurs on nearly all routes on a daily basis, many “route days” have zero values for their volumes in both measurement systems. Thus, it should not be surprising that the median difference is zero.

Differences Between DOIS and CCCS Measured Volumes

| **FSS** | | | |
|---|---|---|---|
| Moments | | | |
| Mean | 2.5 | Std Dev | 38.4 |
| Median | 0.0 | Range | 1,953.0 |
| | | | |
| Quantile | | Estimate | |
| 100% | Max | 1113 | |
| 99% | | 118 | |
| 95% | | 7 | |
| 90% | | 0 | |
| 75% | Q3 | 0 | |
| 50% | Median | 0 | |
| 25% | Q1 | 0 | |
| 10% | | 0 | |
| 5% | | 0 | |
| 1% | | -30 | |
| 0% | Min | -840 | |

| **Sequenced** | | | |
|---|---|---|---|
| Moments | | | |
| Mean | 32.5 | Std Dev | 271.9 |
| Median | 0.0 | Range | 5,110.0 |
| | | | |
| Quantile | | Estimate | |
| 100% | Max | 2596 | |
| 99% | | 958 | |
| 95% | | 532 | |
| 90% | | 331 | |
| 75% | Q3 | 0 | |
| 50% | Median | 0 | |
| 25% | Q1 | 0 | |
| 10% | | -83 | |
| 5% | | -375 | |
| 1% | | -752 | |
| 0% | Min | -2514 | |

While our examination of CCCS and DOIS volume data showed a general correspondence between the two data sets, there are some differences. A key question is whether or not these differences suggest difficulties with using DOIS data in delivery time regressions. One way to validate the use of DOIS data to estimate a street time regression on both the CCCS and DOIS datasets and then compare the results. However, because CCCS data does not exist at the ZIP Code level, this must be done at the route level. Route level regressions show relatively little variability of time with respect to volume because they do not account for the adjustments to volume made at the ZIP Code level. Thus, they are not useful for estimating variabilities but they can be used to compare data sets.

A standard equation for estimating delivery time is a quadratic equation.

$$\begin{aligned} Del.\ Time \ = \ & \beta_0 + \beta_1\, Volume - \beta_2\, Volume^2 + \beta_3\, Poss.\ Deliveries \\ & - \beta_4\, Poss.\ Deliveries^2 - \beta_5\, Volume * Poss.\ Deliveries + \varepsilon \end{aligned}$$

To compare the data sets, we estimated two equations. The first equation regressed street time, using DOIS street hours per route, on CCCS volume data and delivery points. The second equation again used DOIS street hours but substituted DOIS volume data. The regressions were estimated on the 7,845 observations in the comparison data set. The street time regressions for the CCCS data and DOIS data are presented in the next table.

That table shows that the estimated coefficients are quite close to one another in value and that both equations fit the dependent variable equally well (poorly).

**Street Time Regression (in seconds)**

| | CCCS Data | DOIS Data |
|---|---|---|
| Intercept | 9353.48 | 9470.38 |
| Volume | 3.04 | 2.72 |
| Volume^2 | -0.000299 | -0.000243 |
| Poss. Deliveries | 22.46 | 23.18 |
| Poss. Deliveries^2 | -0.011052 | -0.011376 |
| Volume*Poss. Deliveries | -0.001298 | -0.001334 |
| R^2 | 0.140 | 0.136 |

One can test for a significant difference between two regression coefficients with the following formula:

$$Z = \frac{\beta_1 - \beta_2}{\sqrt{s_{\beta_1}^2 + s_{\beta_2}^2}}$$

This test statistic has a t distribution, with a critical value of 1.96. The following table shows no evidence of statistical difference between coefficients as all values for the test statistic are well below two in absolute value. This indicates that using the DOIS volume data generates the same regression coefficients one estimates with the CCCS data.

**Testing for Significant Differences**
**Across Regression Coefficients**

| Variable | t-statistics |
|---|---|
| Intercept | 0.208 |
| Volume | -1.010 |
| Volume^2 | 0.988 |
| Poss. Deliveries | 0.547 |
| Poss. Deliveries^2 | -0.362 |
| Volume*Poss. Deliveries | -0.105 |

The DOIS daily data cover nearly all of the routes and ZIP Codes in the country and provide volume and time data for every delivery day in the year. The data set is extensive, timely, and acceptably complete. A comparison of DOIS daily data with CCCS data shows that the DOIS daily data are reasonably accurate and consistent. The DOIS daily data would appear to hold the potential for estimating street time variabilities.

## E. The Managed Service Point (MSP) Scan Dataset

### 1. Data Set Description

The Managed Service Point (MSP) program uses mobile data collection devices to scan barcodes placed strategically along the city carrier's line of travel. A series of barcodes are located in the office and on the route so that the supervisor can follow a carrier's progress during the day. This allows the supervisor to monitor the consistency of delivery time. A typical day's observation for a route looks like the following hypothetical example:

MSP Location Report

Delivery Unit: Everytown

| Scan Type | Address | Label Location | Scan Time |
|---|---|---|---|
| Hot Case | | Hot Case | 10:30 |
| Depart To Route | | Out Door | 10:41 |
| First Delivery | 123 Main Street | Box Lid | 10:55 |
| Address 1 | 246 Main Street | Box Lid | 11:28 |
| Address 2 | 359 Main Street | Box Lid | 12:23 |
| Address 3 | 678 Main Street | Box Lid | 12:24 |
| Address 4 | 1215 Washington | Box Lid | 12:53 |
| Address 5 | 6060 Jefferson | Box Lid | 14:11 |
| Address 6 | 6200 Jefferson | Business Door Frame | 14:14 |
| Address 7 | 6500 Jefferson | Box Lid | 14:22 |
| Last Delivery | 6700 Jefferson | Box lid | 15:17 |
| Return to Office | | In Door | 15:55 |

### 2. Key Variables

The data set consists of daily observations on when each bar code was scanned on each DOIS route. The MSP data system reports five scans per day that are potentially useful for cost analysis: 1) Hot Case, 2) Depart Route, 3) First Delivery, 4) Last Delivery, and 5) Return to Office. These scans are useful because they provide the potential for segregating travel to and from the route time from the rest of street delivery time.

### 3. Potential Uses

The Delivery Operations Information System (DOIS) reports daily street hours by route. DOIS does not provide separate time estimates for the various activities that carriers perform on the street such as driving to the route, loading the satchel, or delivering parcels. This limits its use to analysis of total street time. However, it might be possible to use MSP scans for each route to separately identify the driving time to and from the route and the delivery time spent between first and last delivery.

### 4. Evaluation of the Usefulness of Data Set for Cost Estimation

While the MSP scan data appeared to hold the potential for refining use of the DOIS data, further examination revealed that it is not a useful source to supplement DOIS data. The principal reason is that the scans for 'First Delivery' and 'Last Delivery' do not necessarily occur at the beginning and end of the delivery activity. The scan for First Delivery represents the first scan point on the route, but that does not necessarily coincide with the actual first delivery on the route. The same holds true for the Last Delivery scan. Carrier supervisors are instructed to make these scanning points as close as possible to their respective names but because customers need to agree to a MSP barcode being placed on their receptacle, the scan points often do not coincide with the actual first and last deliveries on the route.

As a result, the time between the actual first or last delivery points and their corresponding MSP scans may vary widely across routes and use of these scans could lead to inaccurate inference about what is delivery time and what is travel time. For example, our examination found numerous occurrences where over an hour of time elapsed between the ‘Last Delivery’ scan and the ‘Return to Office’ scan. Another factor that contributes to disparity in these times is that carriers may take their lunch break at either the beginning or at the end of the street time portion of their day. This means they are taking lunch after leaving the office and before arriving at the first delivery or after departing the last delivery and before returning to the office.

Because MSP does not identify where and when lunch is taken on the route, it cannot separate out lunch time from delivery time. Note that in the hypothetical example provided above, 55 minutes elapsed between the scan at Address 7 and the scan at Last Delivery despite the fact that they appear to be only a few blocks away. This could occur if the carrier takes lunch after delivering to Address 7.

In sum, MSP is a management tool that gives supervisors knowledge of a carrier’s general location on the route. The information is useful to the Postal Service in managing street time. For example, it allows the supervisor to monitor the consistency of delivery time. MSP data does not provide the necessary information about specific carrier activities to aid in calculating attributable costs by product. Thus, the Postal Service does not recommend that MSP data be used in our efforts to update the CCSTS.

## IV. INVESTIGATING COST POOL FORMATION

The current street time proportions are based upon the results of the 2002 street time study. They are thus based upon a field study in which city carriers used scans to identify the different activities they undertook on the street. The results of that study broke street time into the following categories:

Street Time Proportions Based upon 2002 CCSTS

| | |
|---|---|
| Delivery On Route Sections | 72.3% |
| Parcel/Accountable Delivery | 5.6% |
| Collections From SLB | 0.3% |
| Travel To/From Route | 9.1% |
| Network Travel | 11.4% |
| Relay | 1.4% |

These proportions exclude breaks, lunch, and loading/unloading the vehicle. Note that parcel/accountable delivery time includes any deviation travel time required for delivery. In addition, collections include collection at both Express Mail collection boxes and regular collection boxes.

This section of the report discusses our investigation into using ongoing datasets to update the street time proportions in lieu of launching another expensive field study. To be useful, an ongoing dataset should cover the entire amount of time the carrier spends on the street and should provide the ability to estimate cost pools that are operationally sensible and statistically reliable. Our review of the different ongoing datasets indicates that the Form 3999 dataset holds the most potential for estimating street time proportions. The results of our investigation into its use are discussed in the next section of the report.

### A. Investigating Form 3999 Data for Forming Cost Pools

An alternative approach to using another field study to update the 2002 cost pool proportions is to attempt to find street time proportions based upon the Form 3999 data. As explained above, the Form 3999 dataset consists of one observation for each city carrier route in the country. The data come from the route evaluation process in which the Postal Service collects, among other things, data on the times the carrier spends in the various street activities.

The Form 3999 dataset breaks out street times into a set of activities that are similar to the groupings used in the 2002 CCSTS study. This reflects that the study was designed with operations input to produce cost pools that reflected the activities performed by carriers on the street. The two sets of activities are not identical, however, so it is important to understand the relationship between them. To that end, a proposed mapping between the 2002 groupings and the Form 3999 groupings is given on the next page. While the activity time definitions are not identical, they are sufficiently close to provide a clear mapping.

| Form 3999 Definitions | Current Definitions |
|---|---|
| Sector Segment | Delivery on Route Sections |
| Relay | Relay |
| Travel To<br>Travel From | Travel To/From Route |
| Vehicle Load<br>Vehicle Unload | Considered Office Time in the CRA |
| Travel Within | Network Travel |
| Acct. Delivery<br>Parcel Delivery | Parcel/Accountable |
| Collection | Collections from SLB |
| Deadhead<br>Personal<br>Customer<br>Gas<br>Break<br>Non-Recurring | Break time is excluded. Other non-recurring time are treated as overhead time. |

The following table provides the street time proportions based upon the Form 3999 dataset. As with the 2002 proportions, the Form 3999 proportions exclude breaks, lunch, other nonrecurring time and loading/unloading the vehicle.

Street Time Proportions Based upon Form 3999 Data

| | |
|---|---|
| Sector Segment Delivery Time | 83.4% |
| Parcel/Accountable Delivery Time | 4.3% |
| Collection From SLB Time | 0.2% |
| Travel To/From Route Time | 5.0% |
| Travel Within Time | 3.1% |
| Relay Time | 4.0% |

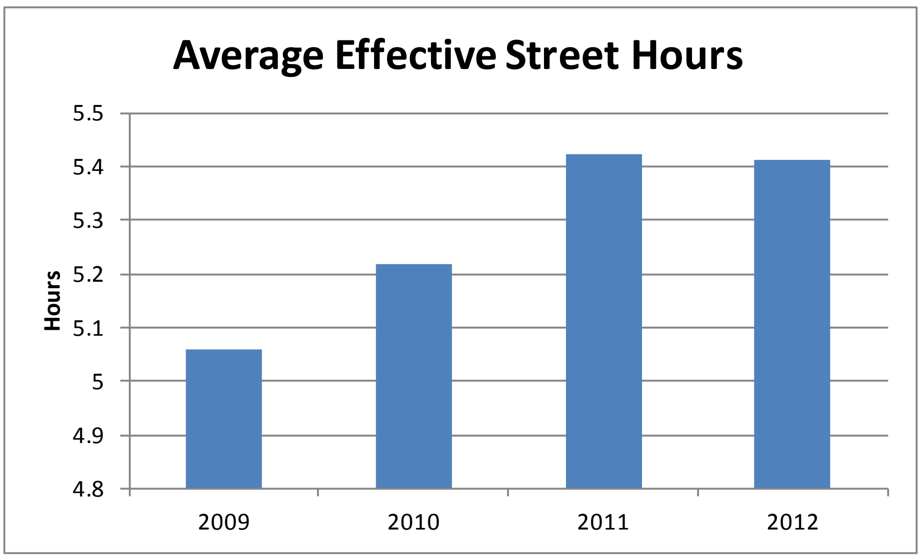
These street time proportions are based upon route evaluation data. Because not all route evaluations are done in the current year and because the route is evaluated prior to its being adjusted, two important questions arise:

- How much do the street time proportions change on a year-by-year basis, thus reducing the utility of older observations?

- How important is it that the data contain pre-reconfiguration route evaluations? Does that mean that the estimated street time proportions are always behind?

Fortunately, both of these issues can be investigated by examining how the street time proportions based on Form 3999 data change though time. This can be done by calculating the street time proportions for each of the years in which there are a substantial number of evaluations. This means street time proportions can be calculated for 2009 through 2012.

We first note that average "effective" street hours per route (excluding breaks, lunch, loading/unloading the vehicle and non-recurring time) have increased by about one-third of an hour since 2009.




The following table presents the street time proportions based on the Form 3999 data for the last four years. Note that this covers a period in which there were a significant number of route consolidations, so it is a useful time period for investigating the two questions listed above.

Street Time Proportions Based on Form 3999 Data

| | 2009 | 2010 | 2011 | 2012 |
|---|---|---|---|---|
| Sector Segment Delivery Time | 83.0% | 83.3% | 83.5% | 83.5% |
| Parcel/Accountable Delivery | 3.7% | 4.0% | 4.3% | 4.8% |
| Relay | 4.2% | 4.0% | 4.1% | 3.7% |
| Travel To/From Route | 5.6% | 5.2% | 4.9% | 5.1% |
| Travel Within | 3.4% | 3.3% | 3.1% | 2.6% |
| Collections From SLB | 0.1% | 0.1% | 0.2% | 0.2% |
| Number of Observations | 3,371 | 24,915 | 96,087 | 16,007 |

The table reveals that the street time proportions have been stable on a year-over-year basis. For example, the time spent in sector segment delivery, the largest cost pool, has stayed relatively constant over the four year period. There has been a modest increase in parcel/accountable delivery time and decreases in both travel to and from time and network travel time, but this reflects the relative growth in parcel volume and the efforts by the Postal Service to reduce vehicle travel time on routes. Moreover, the changes are not large enough to have a material effect on estimated volume variable costs.

In addition, because there were a large number of route reconfigurations during this period, the stability of the street time proportions mitigates the concern about using "pre-reconfiguration data." Although the configuration of an individual route can change substantially after a route evaluation, the overall street time proportions are not greatly changed.

Finally, because of the large amount of data, it would be possible to use the more recent route evaluations to form the cost pools. For example, the following table presents the street time proportions based upon the most current (2011 and 2012) evaluations as well as all evaluations. This comparison shows the results to be stable.

Street Time Proportions Based upon Form 3999 Data

| | Current Route Evaluations | All Route Evaluations |
|---|---|---|
| Sector Segment Delivery Time | 83.5% | 83.4% |
| Parcel/Accountable Delivery | 4.3% | 4.3% |
| Collections From SLB | 0.2% | 0.2% |
| Travel To/From Route | 4.9% | 5.0% |
| Travel Within | 3.0% | 3.1% |
| Relay | 4.0% | 4.0% |
| Number of Observations | 112,094 | 140,501 |

### B. Comparing the Time Proportions across the Two Datasets

The following table presents a comparison of the time proportions derived from the two datasets. Review of the table reveals that there are differences in the sizes of some of the time proportions. Analysis of the differences reveals that they arise for two reasons: (1) Differences in the activity definitions and (2) Restructuring of routes to reduce travel time.

Street Time Proportions from Two Datasets

| Activity | 2002 CCSTS Proportions | Form 3999 Proportions |
|---|---|---|
| Delivery On Route Sections / Sector Segment Delivery | 72.3% | 83.4% |
| Parcel/Accountable Delivery | 5.6% | 4.3% |
| Collections From SLB | 0.3% | 0.2% |
| Travel To/From Route | 9.1% | 5.0% |
| Network Travel / Travel Within | 11.4% | 3.1% |
| Relay | 1.4% | 4.0% |

A first difference occurs in the definitions of Network Travel Time from the CCSTS data and Travel Within Time from the Form 3999 data. The definition of Network Travel captures travel that takes place outside of delivery sections, including all time spent traveling between route sections, or to and from collection boxes. It includes both driving time and walking time. Travel Within in the Form 3999 data, in contrast, captures the amount of time spent moving the vehicle within the delivery sections.

These definitional differences are illustrated in the following table. Note that the Form 3999 data show virtually no Travel Within for foot routes whereas the CCSTS data show a large proportion of time for Network Travel Time. This is because Travel Within, by definition, excludes any walking time, which exists extensively on foot routes. More generally, the table shows that the Form 3999 Travel Within variable is capturing a different part of street activity than Network Travel Time.

**Travel Within and Network Travel Time Proportions by Route Type**

| | Curb | Dismount | Foot | P&L | Other |
|---|---|---|---|---|---|
| 2002 CCSTS Data | 13.2% | 9.1% | 23.2% | 10.1% | 13.6% |
| Form 3999 Data | 1.2% | 2.2% | 0.8% | 4.4% | 2.9% |

This means that Network Travel Time is incorporated into one of the other Form 3999 activities and the data suggest that it is incorporated into Sector Segment Time. This can be seen by comparing the two measures of delivery time from the two data sets, Sector Segment Time and Regular Delivery Time. As the following table shows, the Sector Segment Time is well above Regular Delivery Time because it includes activities like travelling in between route sections.

**Sector Segment and Regular Delivery Time Proportions By Route Type**

| | Curb | Dismount | Foot | P&L | Other |
|---|---|---|---|---|---|
| 2002 CCSTS Data | 70.2% | 73.8% | 54.1% | 74.5% | 67.2% |
| Form 3999 Data | 86.7% | 85.2% | 82.8% | 81.4% | 81.4% |

In fact, when the two time proportions are combined, we see that the Form 3999 combination is just slightly above the 2002 CCSTS combination, which is consistent with Postal Service optimization efforts to reduce travel time relative to delivery time.

**Extended Delivery Time Proportions by Data Base**

| | Curb | Dismount | Foot | P&L | Other |
|---|---|---|---|---|---|
| 2002 CCSTS Data | 83.4% | 82.9% | 77.3% | 84.6% | 80.8% |
| Form 3999 Data | 87.9% | 87.3% | 83.6% | 85.9% | 84.3% |
| Difference | 4.5% | 4.4% | 6.3% | 1.3% | 3.5% |

These definitional differences imply that if Form 3999 data are used to form cost pools, then new variabilities, which are consistent with the Form 3999 definitions, are required for accurate costing.

## C. Summary

Our review of the Form 3999 data suggests that it could be useful for forming street time cost pool proportions. The data are relatively current and cover virtually all of the city carrier routes in the country. The Form 3999 data provide cost pool proportions that comport with operational reality. In fact, as explained to us by city carrier operational experts, the Form 3999 activities reflect the way that Postal Service carrier operations characterize street time activities. Finally, the Form 3999 data provide time proportions that are not erratic through time but show evidence of the evolution of the city carrier network in expected ways.

## V. INVESTIGATING VARIABILITY ESTIMATION

In the past, the variabilities used for each cost pool have come from a variety of sources. Those sources are listed in the following table:

Variability Sources for Current Cost Pools

| | |
|---|---|
| Delivery On Route Sections | Econometric Model |
| Parcel/Accountable Delivery | Econometric Model |
| Collections From SLB | Historical Special Study |
| Travel To/From Route | Overall Street Variability |
| Network Travel | Institutional Cost |
| Relay | Overall Street Variability |

Estimation of appropriate variabilities for street time costs includes resolution of a number of important issues. Those issues are presented and discussed in this section of the report. In addition, it presents what we found out in our review of existing data systems as it applies to variability estimation.

### A. Determining the Variabilities to be Estimated

A first choice is which variability or variabilities to estimate. The simplest model structure is provided by estimating a single overall variability for all street time. For example, the DOIS daily data set records total street hours for city carrier routes across the country and these measured times could be used to estimate the overall variability of street time. Although street time is made up of a number of different cost pools, those individual times are additive, raising the possibility that a single total street time model could provide the same overall volume variable cost, with the consumption of less time and resources, as the overall volume variable cost provided by using the set of the cost-pool specific variabilities. In addition, such an approach would preclude the need to spend resources on estimating cost pool proportions.

To demonstrate this point, consider a simple example in which total street time (ST) is made up of two different cost pools ($ST_1$ and $ST_2$), each with their own variabilities. Suppose that the true cost equations for each of those cost pools have a quadratic form in delivered volume with the following parameters:

$$ST_1 = \alpha_0 + \alpha_1 V - \alpha_2 V^2$$

$$ST_2 = \beta_0 + \beta_1 V - \beta_2 V^2$$

The total volume variable cost would be the sum of the two cost pools' volume variable costs. Each cost pool's volume variable cost is the product of its time proportion ($ST_i/ST$) and its variability:

$$VVC_1 = ST * \left(\frac{ST_1}{ST}\right) * \left[\frac{[\alpha_1 - 2\alpha_2 V]V}{ST_1}\right]$$

$$VVC_2 = ST * \left(\frac{ST_2}{ST}\right) * \left[\frac{[\beta_1 - 2\beta_2 V]V}{ST_2}\right]$$

Now consider estimating a single variability for street time. In this approach, total street time would be specified as a quadratic function of volume:

$$ST = \gamma_0 + \gamma_1 V - \gamma_2 V^2$$

This specification provides an overall amount of volume variable cost:

$$VVC = ST * \left[\frac{[\gamma_1 - 2\gamma_2 V]V}{ST}\right]$$

The key question, then, is identifying the conditions under which the total volume variable cost under the single variability approach, VVC, equals the total volume variable cost, $VVC_1 + VVC_2$, under the cost pool approach. This requires:

$$VVC = VVC_1 + VVC_2$$

. Substituting from the above equations yields:

$$ST * \left[\frac{[\gamma_1 - 2\gamma_2 V]V}{ST}\right] = ST * \left(\frac{ST_1}{ST}\right) * \left[\frac{[\alpha_1 - 2\alpha_2 V]V}{ST_1}\right] + ST * \left(\frac{ST_2}{ST}\right) * \left[\frac{[\beta_1 - 2\beta_2 V]V}{ST_2}\right]$$

Simplifying yields:

$$[\gamma_1 - 2\gamma_2 V] = [\alpha_1 - 2\alpha_2 V] + [\beta_1 - 2\beta_2 V]$$

Or:

$$\gamma_1 - 2\gamma_2 V = (\alpha_1 + \beta_1) - 2(\alpha_2 + \beta_2)V$$

This condition will hold as long as the models are correctly specified and have the same right-hand-side variables.

However, previous research in carrier street time costs has demonstrated not only that street time is made up of a number of different cost pools, but also that those cost pools have different cost drivers. This raises the possibility that an aggregated variability equation could be subject to specification errors and undermines the attractiveness of such an approach.

## B. Determining the Model to Be Estimated

Another important choice in estimating a cost pool's variability is specifying the model in which the variability will be estimated. Model specification has two main aspects, selection of the model's functional form and selection of the variables to be included in the model. Both of these issues will be discussed in this section of the report.

### 1. Selecting the Functional Form

Previous empirical work in carrier street time has determined that carrier street time is subject to economies of density. This means that as additional mail is delivered over the same number of delivery points, the delivery cost per piece falls (although total delivery cost rises). Accordingly, the choice of functional form should allow for this possibility and should let the data determine the degree of density economies. This would rule out, for example, use of a linear functional form.

Research in the area of carrier street time has identified two functional forms that can be successfully used in estimating street time variabilities: the quadratic form and the translog form. Both are “flexible” functional forms in the sense that they impose relatively little restriction on the shape of the delivery time regression equation. Thus, unlike a restrictive functional form like the Cobb-Douglas form, the flexible functional forms allow the data to dictate the shape of the estimated function. This means that the degree (or even presence) of density economies is not pre-ordained by the model selection.

The translog form has a major drawback for estimating street time regressions. Because it is a logarithmic form, the translog cannot be used to estimate equations in which the right-hand-side variables take zero values. Because certain cost drivers, such as FSS mail delivered or sequenced mail delivered, can take on zero values at both the route and ZIP Code levels, the quadratic functional form is more useful for estimating street time regressions.

### 2. <u>Selecting Variables to Include in the Model</u>

When constructing a model to estimate a variability for a certain activity, there are two groups of variables to be selected, the cost drivers and the characteristic variables. The cost drivers are variables which are closely related to both the incurrence of cost in the activity being studied and to the final product, the delivery of mail products to customers. In the carrier street time activity, the cost drivers are the volume of mail delivered and the delivery points to which the mail goes. Characteristic variables influence the relationship between cost drivers and the cost they generate. Characteristic variables can reflect certain aspects of the production technology or can embody constraints on the construction process. In the area of carrier street time, characteristic variables include the nature of the delivery process (e.g. curbline delivery vs. park and loop delivery) and the geographic area over which delivery must take place.

An important consideration when selecting the volume cost driver(s) is to choose between an aggregated approach and a disaggregated approach. In the aggregated approach, delivery time is related to sum of all volume being delivered, whereas in the disaggregated approach, total volume is subset into volume groupings assumed to have the same cost-causing characteristics. The advantage of an aggregated approach is that it eliminates the multicollinearity problem that occurs in street time regressions and increases the likelihood of estimating precise regression coefficients. The disadvantage of the aggregated approach is that is assumes that all volume subaggregates have the same marginal times.

This suggests that the choice between an aggregated and disaggregated approach relates to the relative precision of the estimates in the two methods and the degree to which there are material differences in the marginal times among the disaggregated volume groups. The greater the gain in precision in estimation from aggregation, the greater the advantage of the aggregated approach, and the greater the true differences in marginal times for the disaggregated volume groupings, the greater the disadvantage of the aggregated approach.

To see this, consider the following example. Suppose that there are two volume groupings, $V_1$ and $V_2$ that make up total delivered volume, V. This example has two cases. In the first case, the marginal times are close together and there is a large gain in precision from aggregated estimation. In the second case, the differences in marginal times are much larger and there is less of a precision gain from aggregation.

Examples of the Advantage and Disadvantage from Volume Aggregation

| Variable | $V_1$ | $V_2$ | $V$ |
|---|---|---|---|
| True Marginal Time | 2 | 3 | 2.5 |
| 95% Confidence Interval for the Estimated Coefficient | (1.5 - 2.5) | (1.0 - 5.0) | (2.2 - 2.8) |
| | | | |
| True Marginal Time | 2 | 6 | 4 |
| 95% Confidence Interval for the Estimated Coefficient | (1.5 - 2.5) | (5.5 - 6.5) | (2.5 - 5.5) |

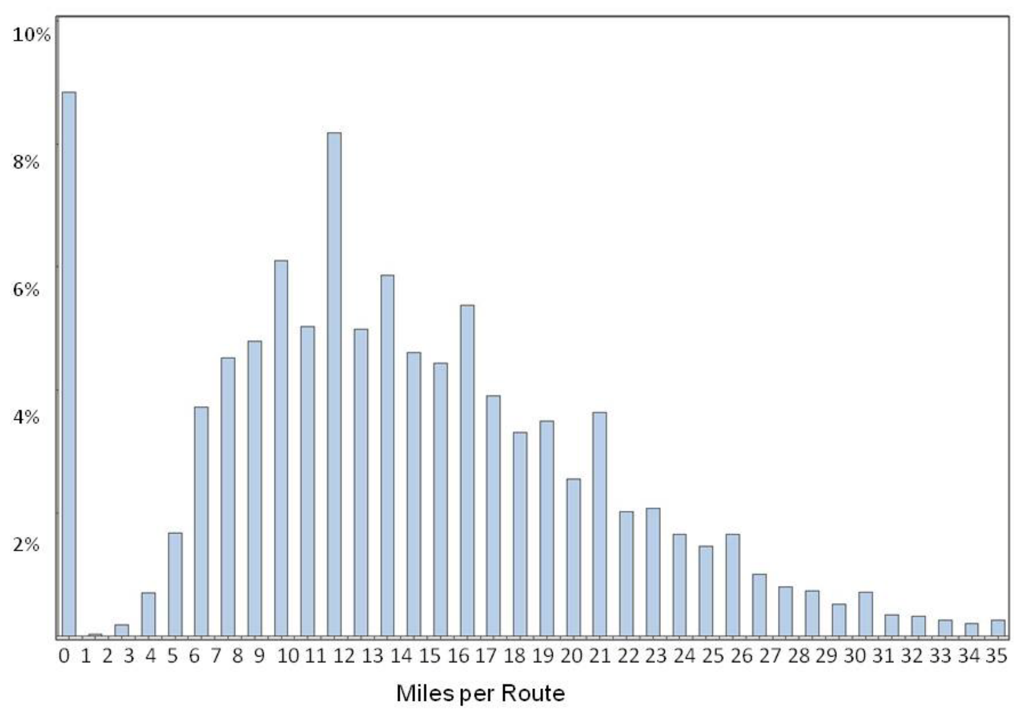
If a disaggregated approach is chosen, then the cost driver groupings must be selected. Ideally, the groupings should capture similar volumes that have the same cost-causing characteristics. For example, suppose that all pieces of DPS mail, regardless of class or rate category, have the same cost causing characteristics on the street. Then it would be appropriate to have a volume grouping for DPS mail. Previous work in carrier street time has shown that there are similar cost-causing characteristics among the mail delivered from the same bundle or container.[3]

[3] There are two primary cost causing characteristics on the street: (1) the likelihood than an additional piece of mail will cause the carrier to stop at a delivery point and thus incur deviation time and (2) the additional time created by handling the mail at that stop.

Given the current state of delivery technology, the following volume groupings are consistent with this approach for a regular delivery equation:  DPS mail, cased mail, FSS mail, and sequenced mail.  In addition, two other possible volume groupings are small parcels not included in cased mail and mail collected from customer receptacles.

In addition to cost drivers, characteristic variables may be required for accurate estimates of street time variabilities.  One of the characteristic variables relates to geography over which the mail is delivered.   A distance measure available at the route level is the base vehicle miles driven on each route.  This information is recorded for each route in the country and averages about 13 miles per route per day.  Not surprisingly, there is a lot of variation in miles driven across the country and there are a very small number of routes that report hundreds of daily miles of routes driven.  However, just 1 percent of the routes have more than 35 daily miles.  Once those routes are eliminated, the distribution looks reasonable.

Distribution of Base Vehicle Miles Per Route



One reason for variation in mileage is the different types of routes. For example, because this measure is for vehicle miles, we would expect it to be virtually zero for foot routes.

Proportions of Routes With Zero Base Vehicle Miles

| | **Curb** | **Dismount** | **Foot** | **P&L** | **Other** |
|---|---|---|---|---|---|
| % Route with Zero Veh. Miles | 1.95% | 2.49% | 99.97% | 4.87% | 68.75% |

In addition, this distance variable can be summed across the routes in a ZIP Code to calculate the total vehicle miles per ZIP Code. However, the focus of this characteristic variable is the extent to which it creates variations in delivery cost across routes or ZIP Codes with the same volume and delivery points. This means that we must be careful to avoid, for example, measuring a “high miles” route or ZIP Code simply because it happens to have many delivery points. To avoid this problem and to capture the effect of geographic density, we calculate a measure of the average distance per delivery point by dividing each route or ZIP Code’s total vehicle miles by its total possible deliveries:

Miles Per Delivery (MPD) = (Vehicle Miles) / (Possible Deliveries)

The other characteristic variable captures the nature of the technology of delivery to the extent it affects delivery time. Each route in the Form 3999 data set is identified by a “delivery code.” The distribution of these codes is presented in the following table:

| Delivery Code | Route Type | # of Routes |
|---|---|---|
| 1 | Curb | 36,007 |
| 2 | Dismount | 23,295 |
| 3 | Foot | 7,824 |
| 4 | Park and Loop | 73,016 |
| 5 | Other | 339 |

** Note that 20 routes did not have a recorded route type.*

However, the main distinction in delivery technology is whether the route includes a substantial amount of walking time. Both foot and park and loop routes include large amounts of walking time and would be expected to require more delivery time for the same amount of volume and delivery points. It is therefore useful to split routes into one of two types, either “motorized” or “walking,” using the following classification:

If Delivery Code = 1 or 2, then its Delivery Type is a “motorized” route.

If Delivery Code = 3, 4, or 5, then its Delivery Type is a “walking” route.

For an individual route, the delivery type variable either takes the value of zero for a motorized route, or a value of one for a walking route. For a ZIP Code, the delivery type variable measures the percentage of routes which are walking routes. Thus, if a ZIP Code has a value of 0.75 for its delivery type then three-quarters of its routes are walking routes.

## C. Determining the Unit of Observation

A next choice to be made is the unit of observation for the estimation of cost pool variabilities. For city carrier street times, there are two logical possibilities, at the level of a route and at the level of a ZIP Code. The appeal of a route-level analysis is that it reflects the basic level at which hours are incurred -- by carriers delivering mail on their routes. However, the Postal Service manages its delivery network at the ZIP Code level, not at the route level. It uses routes primarily as an organizing device to accomplish delivery, as efficiently as possible, given the constraints of its labor agreement and the geography of the delivery area.

While it does not take exactly eight hours to deliver each city carrier's route each day, the Postal Service does adjust routes to the eight hour standard.[4] When volume was increasing, the Postal Service added routes within a ZIP Code, making use of the eight hour standard. Thus, if volume grew substantially on a route, it would be "adjusted" by offloading some of its volume to other, perhaps new, routes as it was moved back to approximately eight hours a day. Thus, it was often the case that the daily total hours in the ZIP Code were rising more rapidly than the daily hours per route in that ZIP Code.

The same process works in reverse as volume falls. When routes fall below the eight hour standard, they are "adjusted" and they acquire volume from other routes which are also being adjusted. This leads to a reduction in routes. In this instance, the

[4] In addition, street time is only part of the eight hour day, which also includes office time. This raises the possibility that there could be more variation in street time with respect to volume despite the eight hour standard. Experience, however, suggests otherwise. In general, office time is more closely linked to volume so a route with growing volume will see its office time expand, which could cause a route adjustment to be required even without a change in street time.

daily total hours in the ZIP Code fall faster than the daily hours per route in the ZIP Code.

Because of this adjustment process, the variability of hours with respect to volume at the route level is attenuated. The route level variability is not necessarily zero, because the adjustment process is not continuous and the eight hour day is a target, not an absolute standard. But the amount of variation in hours on a route with respect to volume will be well below the actual overall variation in hours in the ZIP Code, and an estimated route-based variability will be biased downward.

To estimate the degree of this bias, delivery time variabilities at the route level and the ZIP Code level can be estimated using the same model specification and data. For example, because it contains data for virtually all city carrier routes in the country, the Form 3999 data can be used to estimate a volume variability at both the route level and at the ZIP level. This is because the route level data can be aggregated to the ZIP Code level.

We estimate a quadratic model using the Form 3999 data with “sector segment” or delivery time as the dependent variable and total volume delivered (including DPS mail, cased mail, FSS mail and sequenced mail delivered) and possible deliveries as the cost drivers. We also include the number of miles per delivery point and the delivery type as characteristic variables. The econometric model has the following form:

$$\begin{aligned} \begin{matrix} Delivery \\ Time_i \end{matrix} = {} & \beta_0 + \beta_1 V_i + \beta_2 V_i^2 + \beta_3 PD_i + \beta_4 PD_i^2 + \beta_5 V_i * PD_i + \beta_6 \left(\frac{Miles_i}{PD_i}\right) \\ & + \beta_7 \left(\frac{Miles_i}{PD_i}\right)^2 + \beta_8 DT_i + \beta_9 DT_i^2 + \varepsilon_i \end{aligned}$$

The model is estimated on both 140,331 route observations and the corresponding 10,642 ZIP Code level observations.

The following table presents the results for the route-level data.[5] The table shows that the cost drivers have their expected signs (positive linear terms and negative quadratic terms). Miles per delivery point have little explanatory power at the route level as both geographically dense and geographically disparate routes have approximate eight hour days. In general, the explanatory power of the equation is small and the estimated variability of 11 percent is low.

Delivery Time (Hours) Equation - Form 3999 Data

Route Level Data

| Variable | Coefficient | t-statistic |
|---|---|---|
| Intercept | 1.87 | 14.10 |
| Volume | 4.624E-04 | 9.49 |
| Volume Sq. | -5.912E-09 | -2.72 |
| Poss. Deliveries | 4.540E-03 | 17.20 |
| Poss. Deliveries Sq. | -1.720E-06 | -11.10 |
| Volume* Poss. Deliveries | -3.397E-07 | -4.90 |
| Miles Per PD | 0.4982 | 1.10 |
| Miles Per PD Sq. | -0.0262 | -1.16 |
| Delivery Type | 0.0487 | 1.38 |
| | | |
| # of Observations | | 140,331 |
| R^2 | | 0.005 |
| Volume Variability | | 11.4% |

[5] Because the delivery type variable is discrete at the route level, it can only be included linearly in the equation.

The next table presents the results from estimating the model at the ZIP Code level. Again the cost drivers have the expected signs but now the characteristic variables play an important role in explaining variations in delivery time. The time to deliver a given amount of volume is higher in ZIP Codes that are geographically disparate and it is also higher in ZIP Codes that have a high proportion of walking routes. Moreover, the explanatory power of this equation is much higher than at the route level. The estimated variability of approximately 40 percent is in the same range as the overall variability of delivery time estimated with CCSTS data.

Delivery Time (Hours) Equation - Form 3999 Data

ZIP Code Level Data

| Variable | Coefficient | t-statistic |
|---|---:|---:|
| Intercept | -14.88 | -16.83 |
| Volume | 9.082E-04 | 17.92 |
| Volume Sq. | -2.130E-09 | -3.34 |
| Poss. Deliveries | 4.940E-03 | 22.10 |
| Poss. Deliveries Sq. | -7.376E-08 | -4.87 |
| Volume* Poss. Deliveries | 1.037E-08 | 1.81 |
| Miles Per PD | 214.40 | 11.53 |
| Miles Per  PD Sq. | -95.36 | -9.49 |
| Delivery Type | 34.35 | 12.98 |
| Delivery Type Sq. | -21.65 | -8.48 |
| # of Observations | | 10,624 |
| R^2 | | 0.762 |
| Volume Variability | | 39.9% |

A comparison of the route level results with the ZIP Code level results produces evidence of attenuation at the route level. The overall explanatory power of the route-level equation is much lower than the ZIP Code level equation and the estimated variability is substantially smaller.

## D. Summary

We investigated a number of important issues relevant for estimating variabilities for street time activities and identified a number of results. First, our analysis suggests that it would be difficult for a top-down approach that estimates a single variability for all of street time to replicate the volume variable costs calculated through the cost pool approach. Second, we identified conditions under which an aggregated volume approach might provide preferred results compared to a disaggregated volume approach. These conditions are empirical and depend upon the relative difference in product grouping marginal times and the degree of imprecision caused by including many volume groupings.

Based upon the operational changes since the last street time study, we determined that appropriate volume groupings for delivered mail include DPS mail, all cased mail, FSS mail and sequenced mail. We also identified two characteristic variables, a measure of geographic dispersal and a measure of delivery type that would be useful to help explain variations in street time. Next, we investigated whether a delivery time variability should be estimated at the route level or ZIP Code level both from an operational perspective and from an empirical perspective. We concluded that it is likely that estimation of a variability at the route level would be biased downward, and produced empirical evidence to support that conclusion.

Finally, we investigated the possibility of using Form 3999 data to estimate delivery time variability. This approach has several advantages:

- The Form 3999 data would save the time and expense of undertaking a field study.
- The Form 3999 data could be used to estimate either an aggregated or disaggregated variability.
- The Form 3999 data include a variable called “sector segment” time, which is akin to regular delivery time in the current CRA cost model.
- Using Form 3999 data would provide variabilities that match a set of cost pools also based upon Form 3999 data.
- The Form 3999 data include separate measures for DPS volume, cased volume, FSS volume and sequenced volume.
- The Form 3999 data can be matched with data that can provide route type and vehicle mileage information which is useful for constructing characteristic variables.

However, the use of Form 3999 data raises a number of concerns:

- It does not include a measure of volumes collected at customer delivery points.
- It includes some parcels in with cased flats.
- It does not include volumes for accountables.

We also investigated empirically estimating a variability for the parcel/accountable delivery cost pool and concluded that the estimation of a parcel/accountable variability is likely to require a special field study. This conclusion is based upon a number of reasons.  First, the Form 3999 data does not include the volume of accountables delivered.  Next, although the Form 3999 data does include a measure of parcels delivered, but it is not clear that measure corresponds to deviation parcels.

## VI. CONCLUSIONS

In this scoping study, we reviewed important changes in the delivery environment that have occurred since the 2002 City Carrier Street Time Study (CCSTS), investigated ongoing databases to discover their potential usefulness in estimating cost pools and variabilities, explored model specification choices including options for both cost pools and estimating variabilities. Our research revealed, among others, the following important points:

- There have been a number of important changes in operations that should be considered when pursuing street time cost analyses. These changes include things like revision in the bundles or containers carriers use on the street, the introduction of new products and technologies, and changes in the way city carriers are managed.

- The Carrier Optimal Routing System (COR) and the Managed Service Point (MSP) scan program both provide Postal Service management with useful tool for better configuring and managing the city carrier network. They do not appear, however, to provide data useful for estimating street time costs.

- The Form 3999 Dataset holds potential for updating the street time proportions. Time proportions based upon the Form 3999 data are generally consistent with the current approach and are stable through time.

- Variability models with a single aggregated volume variable provide consistent results across data sets.  However, this approach requires the assumption that the marginal times for delivery are the same across all shapes.

- None of the ongoing datasets include data on collection of mail from customers' receptacles or accountables delivered.

- Operational and empirical evidence indicates that use of route-level econometric models lead to understated variabilities..

- A comparison of the DOIS Daily data with CCCS data reveals that the DOIS data are generally consistent with the CCCS data and are potentially useful for estimating econometric models used for calculating variabilities.

- Based upon the operational changes since the last street time study, we determined that appropriate volume groupings for delivered mail include DPS mail, all cased mail, FSS mail and sequenced mail.  We also identified two characteristic variables, a measure of geographic dispersal and a measure of delivery type that would be useful to help explain variations in street time.

- The Form 3999 data hold potential for estimating a street or delivery time variability.  Such an approach holds a number of advantages but also raises a number of concerns.

- It appears that that a special study will be needed to estimate deviation parcel and accountable variabilities.